BI 311, Biostatistics

Location: Henry Hall Lab 1
Meeting times: MWF, 8:30 – 9:20AM
Instructor: Michael Dohm, PhD
Office: Henry Hall 6; Office Hours Mon 10:30 – 11:20 AM; By appointment
E-mail: mdohm@chaminade.edu

Course Description

Biostatistics is a hands-on course designed to provide students with the opportunity to develop analytical and statistical reasoning skills appropriate for biology. Statistical reasoning may be defined as the ability to understand and use probability arguments and to recognize the distinction between specific and generalized conclusions. After an introduction to data sets and the three philosophical positions of statistical hypotheses (frequentist, likelihood, and Bayesian), students will learn to apply statistical reasoning to questions about biological processes. Students will learn about error concepts and to identify and estimate via experimental design or simulation impacts of experimental versus biological sources of error on conclusions. Students will move from descriptive statistics and issues of parameter estimation to factorial experimental design (ANOVA, contingency tables), and general linear and logit models, all with examples and problem sets from biology.

Learning Outcomes

During the semester, students will be given an introduction to (1) statistical tools and approaches used in biological and biomedical research, and (2) experimental design and analysis. At the conclusion of the course, students will be able to demonstrate:

1. Knowledge of data presentation through use of summary tables, bar charts, and scatter plots.
2. Ability to identify and distinguish between data types, variables, and parameters and provide examples from biology.
3. The ability to choose among kinds of statistical tests based on data types.
4. Knowledge of the difference between description and hypothesis testing.
5. The ability to present experimental design and data gathering approaches to test hypotheses on a problem in biology.

Statistical reasoning

Statistical reasoning may be defined as the ability to understand and use numbers to communicate findings and to support opinions. Statistical reasoning is important to your future, for two reasons. First, the biological sciences are, by their nature, an experimental science: all disciplines in biology involve the presentation and analysis of data. Consider a biomedical example: You work as an AIDS counselor in a major U.S. city. Before you is the result of a male client's HIV test. It is positive, meaning the ELISA and Western Blot tests detected HIV in the blood sample. The client, who insists that he does not fit any known risk behavior group, now sits before you, waiting to hear the results. Your training has given you the following facts: (1) less than 0.01% of men fitting this category (heterosexual, no IV drug use) are infected with HIV in the U.S.; (2) the sensitivity of the tests is very high, 99.9% (sensitivity is defined as the percentage of individuals with a disease who are correctly classified by a test as having the disease); and the specificity of the tests is also high, 99.9% (specificity is defined as the percentage of individuals without a disease who are correctly classified by a test as not having the disease). How do you communicate the client's test results?

Second, statistical reasoning is common to our daily lives. The above example illustrates the problem from the perspective of a counselor, but as citizens, we are faced with a barrage of numbers: our risk of developing a particular disease (breast cancer, 1 in 9 women; prostate cancer, 1 in 9 men, etc.); our risk from dying under general anesthesia (on average, about 0.01% or 1 death every 10,000 U.S. cases); our risk of dying in a car crash versus our risk of dying in a plane crash (see Lecture 1). Data

summaries, projections and predictions, and declarations of probability and likelihood are important tools for communicating complex information. The media commonly employ these tools to present information to you: The latest results from clinical trials of an HIV drug therapy, putative links between diet and health, effects of environmental toxins on health... the list is long. Often, the portrayal of medical and health news in the media is conflicting, and the lack of consistency can be explained in part by failure to communicate the statistical issues. One result from the media's poor presentation of science results is the impression that just about everything and anything can be shown to adversely affect health... not a very useful conclusion. More importantly, exaggerated claims are made with numbers, and it may be difficult to sort fact from interpretation from fiction when numbers are used.

Required Text(s)

1. Your Moodle course website
2. Norman, G. R, and D. L. Streiner. 2008. *Biostatistics: The Bare Essentials*, 3rd edition. People's Medical Publishing House. (ISBN 978-1-550093476) Link to Amazon.com

Other text books (recommended but not required).

1. Everitt, B. S., and T. Hothorn. 2009. *A Handbook of Statistical Analysis Using R*, 2nd edition. Chapman and Hall/CRC. (ISBN 978-1420079333) Link to Amazon.com
2. Glover, T., and K. Mitchell. 2008. An Introduction to Biostatistics, 2nd edition. Waveland Press (ISBN 1577665805). *BI311 was formerly taught as "BI460" in previous Chaminade Catalogs. If you have the required text from BI460, you will be able to use this textbook. Note however that the two texts use different examples, and that the REQUIRED textbook is the one listed above.*
3. Zar, J. 1999. *Biostatistical Analysis*, 4th edition, Prentice Hall Press (ISBN 0131008463).
4. Sokal, R. R., and F. J. Rohlf. 1996. *Biometry*, 3rd edition, W. H. Freeman (ISBN 0716724111).
5. StatSoft, Inc. 2010. *Electronic Statistics Textbook*. Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/ .

Assignments

Homework: There will be up to eight (8) short assignments that will help you understand concepts that we cover in class. Assignments include analyzing data relevant to lecture topics and writing short reports that include presenting experimental designs, methods and statistical results using tables and graphs. You may work together on homework, but each student must turn in their own homework. Your textbook also has many nice problems to work on: I will recommend many, but will not grade work from the text book. Of course, the more you do, the better your understanding will be!

Project: The student will design a project of interest in the biological sciences. The project must include the following elements: experimental design, data collection, data analysis, and conclusions. There are four graded elements:

Proposal: A brief description of your project, hypotheses, and how you will approach testing the hypotheses. You will turn this in twice: a draft, then a final version which adopts and addresses my suggestions.

Experimental Design: Detailed presentation of kinds of observations to be made, how they will be made, treatment groups, sample size (power), sampling scheme, statistical model. You will turn this in twice: a draft, then a final version which adopts and addresses my suggestions.

Oral Report: The student will provide an oral report with graphics (e.g., Powerpoint) to share with the class results of the project. An important component of the report will be responding to questions about experimental design and analytical choices.

Final Written Report: The student will submit a final, written manuscript that follows standard scientific journal format. The report is done in collaboration with the student's group, but each student must turn in their own paper, written only by the student. Specific grading criteria, requirements, and additional details about the projects will be provided as we progress through the semester.

Exams : Three exams:

Exam I will cover material from weeks 1-5.

Exam II will cover material from weeks 6-10.

A Final Exam, noncomprehensive (=Exam III), covering material from weeks 11-15.

The exams will focus on your understanding of important concepts, your ability to evaluate experimental designs, and your ability to recognize appropriate statistics to be employed given sets of data or particular experimental designs.

Class Participation : You are expected to come to class prepared, to participate fully in class by asking questions, bringing attention to media announcements of relevance to biostatistics, by helping in data collection needed for homework projects. I also encourage all of you to assist fellow students with computer and statistical software-related questions.

Grading

Item	**How many?**	How often may I expect this item?	**How many points is each item worth?**	**Total points from this item towards my final grade**
Homework	8	every 1 – 2 weeks	12.5	100
Exams	3	every 5 weeks	100	300
Projects				100
Proposal			10	
Experimental design			10	
Oral report			20	
Written report			60	
Total				**500**

Final grade : Your letter grade will be based on the following point distribution.

Points earned	**Percent of total**	**Letter grade**
450 – 500	90 – 100%	A
400 – 449	80 – 89%	B
350 – 399	70 – 79%	C
300 – 349	60 – 69%	D
≤ 299	≤ 60%	F

Bonus points -- Yes, there will be opportunities to earn a few bonus points from time to time up to a total of 1% of total available points, at my discretion.

Reminders and notices:

1. This computer classroom was designed and is now maintained by Chaminade's Department of Information Technology and the Division of Natural Sciences and Mathematics so that you would have access to a state-of-the art academic computing environment. Money is simply not available to repair abused or stolen computers. Therefore, each person has the responsibility to use the computers responsibly. By using these facilities, you agree to abide by the Computer Room Policies posted in the classroom. These rules include, but are not limited to

- No food or drink
- You agree to adhere to the account setup procedures and use restrictions
- No alterations of software or hardware configurations
- No use of the computers for personal or commercial activities, (except where such activities are otherwise permitted or authorized under applicable University policies)

2. We will meet regularly throughout the semester; please keep me informed of changes in your schedule. The objective of our discussion is to provide the needed context to remove barriers to your understanding of the material. We will also be using a statistics software package (R) that will be intimidating at first -- going it alone is not recommended.
3. Use of music devices and cell phones is prohibited during all Natural Science and Mathematics classes at Chaminade, unless specifically permitted by your instructor. Use of cellphones and music devices in laboratories is a safety issue. In addition, use of cellphones and music devices in any class is discourteous and may lead to suspicion of academic misconduct. Students who cannot comply with this rule will be asked to leave class and may be subject to laboratory safety violation fines. Please refer any questions to the Dean of Natural Sciences and Mathematics.
4. No make up quiz or exam will be granted in the event of an absence. If a student cannot attend a class in which a quiz has been scheduled, the student must notify the instructor no later than the class prior to the scheduled quiz. For example, if a quiz is scheduled for Thursday, then student must approach and receive permission for the absence no later than the prior Tuesday class. In the event of illness, a Doctor's note will be expected.
5. You are encouraged to work together; however, all graded material must be your own. You are also expected to have read and to abide by the "Student Rules of Conduct" which are available in your copy of Chaminade University's Student Handbook.
6. Students with special needs who meet criteria for the Americans with Disabilities Act (ADA) provisions must provide written documentation of the need for accommodations from the CUH Counseling Center (Dr. June Yasuhara; phone 735-4845) by the end of week three of the class, in order for the instructor to plan accordingly. Failure to provide written documentation will prevent your instructor from making the necessary accommodations. Please refer any questions to the Dean of Students and review the procedures at http://www.chaminade.edu/student_life/sss/counseling_services.php.

Tentative* schedule of topics and assignments

Dates	Topics and assignments	Readings
Week 1 16 - 20 Jan	**1/16: ML King Day – No class** **Lecture**: Why statistics is important in biology. Data, samples, populations, and an introduction to probability. Thinking about experiments in biology. How do interesting questions move between the lab, the field, the stages of analysis, towards an understanding of biology? **Computer exercises**: Use of computers to manage research projects (databases, spreadsheets, statistical packages); resources available (free software, help with particular packages, shortcuts). **Projects**: Groups assigned for projects. Schedule group meetings with instructor to propose project	Ch1 – 3
Week 2 23 – 27 Jan	**Lecture**: Data types, descriptive statistics, introduction to sampling, populations, and to probability. **Computer exercises**: Introduction to R Statistical Language; Summary statistics and graphical display. **Assignments**: Homework 1 assigned. **Projects**: Draft of Project Proposals due by Wednesday, 5 PM	Ch1 –3
Week 3 30 Jan - 3Feb	**Lecture**: Normal distribution and Probability. Statistical inference: Hypothesis testing, error rates, with examples of how risk analysis instructs decision-based biology research (e.g., how to interpret results of a diagnostic test in the Doctor's office; how to interpret outcomes from a fence-exclusion project in ecological restoration) **Computer exercise**: Normal distribution and proportions of a curve; sampling distributions **Assignments**: Homework 1 due; Homework 2 assigned. **Projects**: Meet with instructor this week.	Ch4 – 6
Week 4 6 – 10 Feb	**Lecture**: Hypothesis testing, error rates, continued. Comparing two groups (t-tests), comparing more than two groups (ANOVA) **Computer exercise**: one and two sample t-tests; testing assumptions; one-way ANOVA; **Project**: Proposal rewrites due by 10AM Tuesday; Draft project implementation plan due by Thursday, 5PM; Schedule meeting with instructor over proposed project implementation **Assignment**: Homework 2 due; Homework 3 assigned	Ch7 – 8
Week 5 9/20 9/22	**Lecture**: One-way ANOVA, posthoc tests; relationship between ANOVA and regression (GLM) **Computer exercise**: one and two sample t-tests; testing assumptions; Wilcoxon, Mann-Whitney tests. **Project**: Project implementation plan due Tuesday; Data collection should be started no later than this week **Assignment**: Homework 3 due; Homework 4 assigned	Ch8 – 11
Week 6 9/27 9/29	**Lecture**: parametric one and two sample tests, T-tests as special cases of a general approach (GLM); Nonparametric alternatives and additional discussion of statistical assumptions, violations of assumptions, and how they affect conclusions of statistical tests. **Project**: Data collection in progress; **Assignment**: Homework 4 due start of class Thursday	Ch8 – 9
Week 7 10/4 10/6	**Review** in class for Exam 1 **Project**: Data collection continues; Schedule meeting with instructor over data collection progress. **Exam 1**: Thursday	
Week 8 10/11 10/13	**Lecture**: two-way crossed, balanced ANOVA, other two way ANOVA designs; nested, repeated measures; split plot. **Computer exercise**: general linear model (GLM) **Assignment**: Homework 5 assigned	Ch10 – 11
Week 9 10/18 10/20	**Lecture**: two-way crossed, balanced ANOVA, other two way ANOVA designs; nested, repeated measures; split plot. **Computer exercise**: general linear model (GLM)	Ch10 – 11

Dates	Topics and assignments	Readings
	Assignment: Homework 5 due; Homework 6 assigned	
Week 10 10/25 10/27	**Lecture**: Correlation, simple linear regression, fit statistics, Multiple linear regression. Model building **Computer exercise**: GLM **Project**: Data collection completed no later than 1 Nov.; Preliminary project data analysis in progress by 1 Nov.; Draft of manuscript outline due 3 Nov **Assignment**: Homework 6 due; Homework 7 assigned	Ch13 – 14, 16
Week 11 11/1 11/3	**Lecture**: Goodness of fit and contingency table chi-square (crosstabs); **Computer exercise**: Chi-square analyses **Special topic**: Report write-ups: Scientific writing – How to present results of experiments in papers and talks. **Projects**: Meet with instructor re: data collection, description **Assignment**: Homework 7 due; Homework 8 assigned	Ch21 – 24
Week 12 11/8 11/10	No class this week, instructor at ABRCMS conference. **Assignment**: Homework 8 due **Projects: Draft of Materials and Methods due by 5PM, 10 Nov**	
Week 13 11/15 11/17	**Review** in class for Exam 2 **Project**: Preliminary project data analysis Project data collection continues. **Exam 2** in class, Thursday	
Week 14 11/22 11/24	**Project:** Oral reports & presentations Tuesday No class 24 Nov, Thanksgiving break	
Week 15 11/29 12/1	**Project**: Tuesday, Written report due by 5 PM; Thursday, Review and course wrap-up	
Week 16 12/6	8:30 – 10:30 AM Final Exam	